The Evolution of Horizontally Transferred Genes: A Model for Prokaryotes

5

Iñaki Comas and Fernando González-Candelas

Abstract

Horizontal gene transfer is a pervasive evolutionary process which has developed and is still developing an essential role in shaping biodiversity through providing opportunities for innovation, moving determinants of functions among taxa, opening opportunities for colonization of new niches, or acting as a catalyst for adaptation. However, its importance in evolution has only recently begun to be recognized. Reasons for this relatively late recognition of HGT and its relevance stem from two main sources. One is the availability, or rather the lack thereof, of appropriate information to infer the existence of HGT. This shortcoming has now been widely overcome with the large, and still growing at a very fast rate, number of completely sequenced genomes, which allow for precise phylogenetic reconstructions, along with the deployment of theoretical models necessary for inferring evolutionary events in deep-time. The second source is the appreciation that HGT does not leave an indelible stamp on the transferred genes because these continue evolving under a different genomic, and often ecological, environment which usually act synergistically to erase the initially clear marks that go along with those genes. In this chapter we provide an overview of these processes along with a model proposal that will help to better understand the consequences of the continuous evolution of laterally transferred genes at different evolutionary time-scales.

A model from Xanthomonadales and *Legionella pneumophila*

Xanthomonadales are plant pathogens included in the gamma-Proteobacteria division. One relevant feature of Xanthomonadales is that their genomes are the result of multiple events of HGT from other Proteobacteria to the ancestral genome of the group (Fig. 5.1A) (Comas *et al.*, 2006). These results and recent findings by other groups have prompted us to delineate a model for the effects of lateral gene transfer along time on gene contents and phylogenomic analyses of genomes involved in HGT. This model is expected to hold for highly promiscuous bacteria where the main constraints to transfers are those derived from incompatibilities in genomic architecture and, therefore, from the time since divergence of the two lineages considered. This results in web-like structures for the phylogenies of HGT-prone taxa, such as Xanthomonadales, apparently due to absence of selective pressure against the majority of transfers, thus representing the most extreme case of genetic promiscuity among bacteria. This almost free-exchange scenario is represented in Fig. 5.2A.

The model developed for Xanthomonadales and similar taxa in terms of HGT is based on two assumptions:

Date: 09:20 Wednesday 18 April 2012
File: Horizontal Gene Transfer 1P

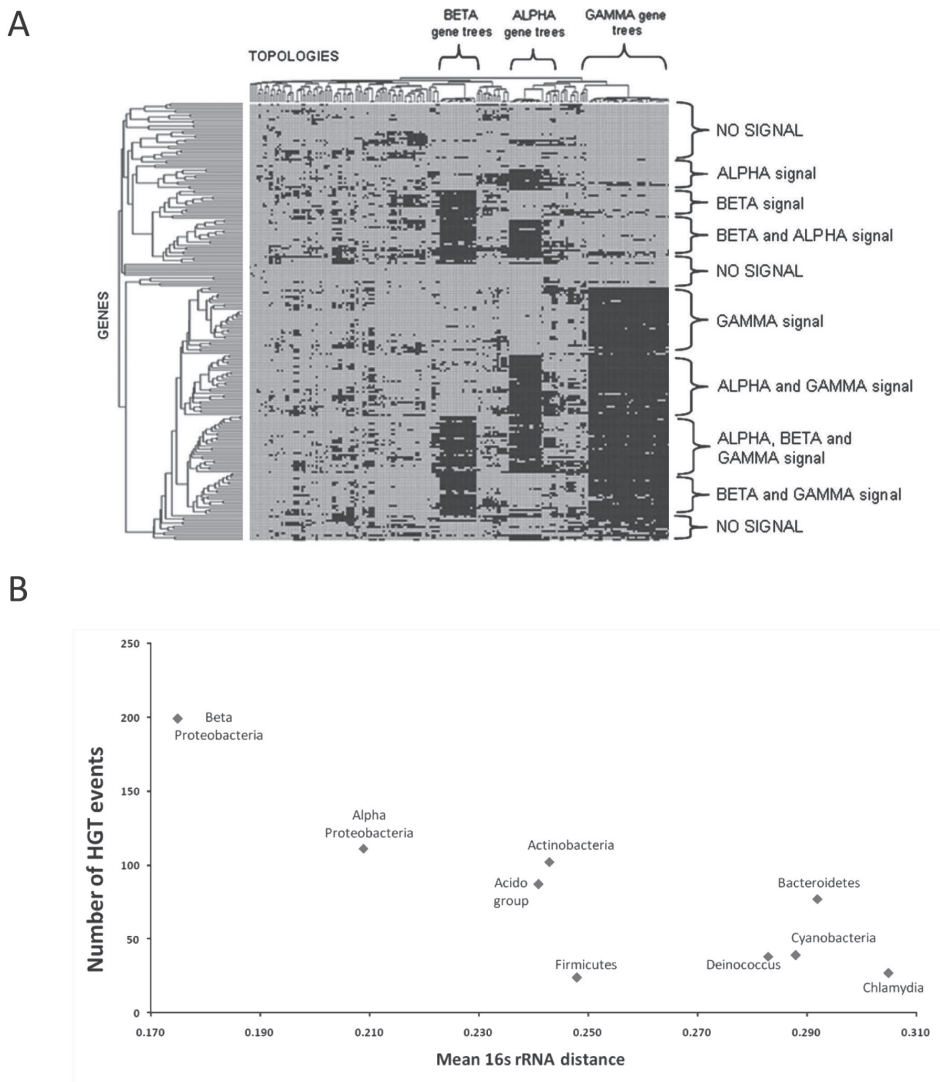
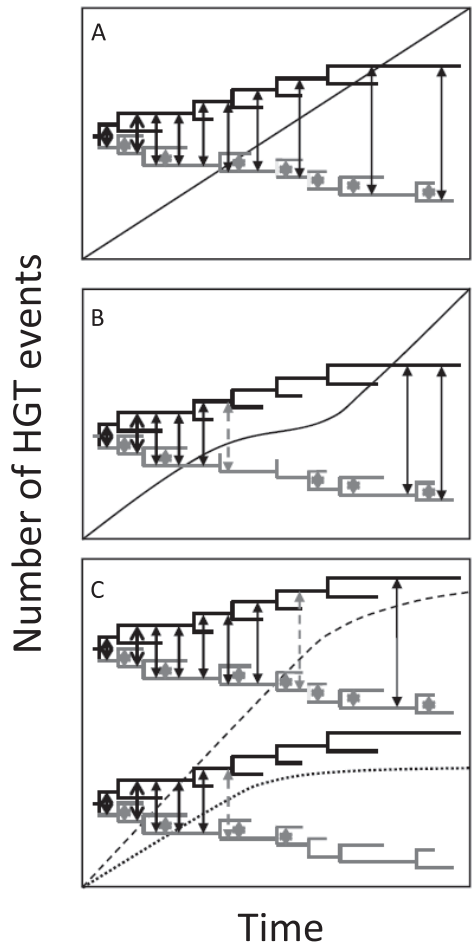


Figure 5.1 Likelihood of horizontal gene transfer decays with phylogentic distance. (A) To differentiate between similar gene tree topologies we computed a tree likelihood comparison statistic that allows discerning between statistically different topologies. The heat-map shown in the figure represents an unsupervised clustering of gene trees and gene alignments according to whether the latter reject (grey dots) or not other gene trees topologies (black dots). Islands of similar gene tree topologies are discovered in this way. As shown in the figure some gene trees place Xanthomonadales as Gammaproteobacteria while others shows a beta or alpha signal, revealing a mosaic genome created by a large number of horizontal gene transfer events. (B) Decay in the number of horizontal gene transfer events detected after analysing the pan-genome of *Legionella pneumophila*. While decay in intra-specific rates of recombination has been known for a very long time, this is the first time that an analogue phenomenon but at much broader phylogenetic distances is proven.



<AU: All arrows are of a very similar width. Is there a better way of representing likelihood? Could the arrows be made thicker to emphasise differences in likelihood? Also, should these graphs have legends?>

Figure 5.2 Changes in the frequency of HGT under different theoretical ecological scenarios illustrated by real examples. (A) The scenario for highly promiscuous bacteria like Xanthomonadales. A continuous import of foreign DNA happened during evolutionary history. This is a free-exchange scenario that is common to other groups and in which gene loss is expected to keep genome sizes on the range observed for bacteria. Many of these gene losses can be the result of natural selection acting on the newly acquired genes. This scenario is likely to happen in free-living or opportunistic bacteria with plenty of opportunities for gene exchange. (B) The scenario for a bacterial lineage in which the income of transferred genes is reduced as it adapts to a new niche thanks to previously acquired genomic islands. Once established, the new lineage starts to diversify and to receive again foreign genetic material. This case is exemplified by the diversification of *Phroclorococcus* strains in ecotypes that differ in the composition and distribution of several genomic islands. (C) The scenario for endosymbionts and some intracellular pathogens (for example *Mycobacterium leprae*). In this case the acquisition of genetic material can result in the opportunity to invade a new host, eventually as a pathogen. Endosymbionts and intracellular pathogens are an extreme example for this case, as their intracellular habitat prevents recombination events and transfer of genes in the majority of cases (dotted line). Non-intracellular genomes retain the capacity to accept foreign genetic material but usually with fewer opportunities (dashed line). The arrows within genealogies represent likelihood of transfers between and within nascent grey and black lineages over time, with likelihoods represented by the thickness of the arrows.

1 Genetic exchange in bacteria is widespread but not random, being more likely between closely related taxa than between evolutionarily distant groups (Gogarten *et al.*, 2002).
2 The consequences of gene transfers on genome phylogenies differ depending on the time elapsed since the transfer and from the time of divergence between the donor and the recipient.

If both principles are true then it is expected that the present preferential partners of Xanthomonadales are other genomes of the same clade, whereas in the past, when Proteobacteria and Xanthomonadales were nascent lineages, the preferential partners were other Proteobacteria, whose extant descendants may now be highly divergent from the Xanthomonadales. Our analysis of Xanthomonadales genomes corroborates this prediction.

Genome-wide phylogenies show two relevant facts: on the one hand, the Xanthomonadales clade is monophyletic for all the gene trees, thus showing that every gene analysed was already present in the last common ancestor of this group. On the other hand, a consensus tree failed to identify a single, common origin for these genes (Comas *et al.*, 2006). The second observation clearly points to ancient HGT to the ancestor of the genomes from other Proteobacteria, a possibility corroborated in further analyses. The first observation is indicating either that vertical inheritance was the dominant force during the diversification of the clade or that there are HGT events not detected by phylogenetic analyses at the genome-level whose signal is also vertical-like. An analysis of the composition of the *X. citri* genome revealed the presence of a significant fraction of atypical genes (22%). Most of them were not detected in our phylogenetic analyses, thus corroborating that recent intra-clade exchanges exist and, therefore, that the cohesion of the clade observed in the summaries of gene trees is fuelled not only by vertical descent but also by horizontal gene transfer as predicted elsewhere (Gogarten *et al.*, 2002).

Let us now consider another model organism we have been working on in our laboratory. *Legionella pneumophila* is the causative agent of legionnaires' disease. It can be found in aquatic environments, free-living or in biofilms, or infecting human macrophages. One remarkable feature of *Legionella* is that it can survive also within amoebas, sharing this niche with many other bacterial species (Coscollá *et al.*, 2011). In fact, the amoeba host has been proposed as a training field for many human pathogens able to invade cells (Molmeret *et al.*, 2005). We have recently performed a phylogenomic analysis of the core genome of four *Legionella pneumophila* strains in search of possible HGT events. This analysis revealed that *Legionella* genomes harbour a large fraction (about 42% of 1700 genes included in the study) of seemingly horizontally transferred genes from different species, most of them with close relatives infecting amoebas. Furthermore, despite these increased chances for HGT deriving from the shared intracellular environment (see also Chapter 4), we were able to demonstrate that the number of HGT events in the *Legionella* genomes decreases with increasing phylogenetic distances from the putative donor group (Fig. 5.1B).

Therefore both the Xanthomonadales and the *Legionella pneumophila* genomes point towards the two main factors regulating the amount of HGT between two species: the phylogenetic distance between them (i.e. intrinsic factors) and the ecological opportunity for the exchange (i.e. extrinsic factors).

Non-random genetic exchanges: internal and external factors

As previously mentioned, there are different factors limiting a widespread, random exchange of genetic material among bacteria. Most of these factors are known but those belonging to the genome architecture/compatibility class have been reported only recently.

Recombination in bacterial populations is known to decrease with divergence. In fact, this is a by-product of the mechanisms involved in the process. The mismatch-correction systems of bacteria are capable of distinguishing foreign DNA based on its genetic divergence from the recipient (Majewski and Cohan, 1998; Cohan, 2002). Therefore, this mechanism represents a barrier to homologous recombination at the population level and, in consequence, it delineates a first boundary to gene exchange among lineages.

For years, bacteria were considered as clonal both at the population and the species levels (Orskov and Orskov, 1983; Stenderup and Orskov, 1983). MLST analyses of different bacterial populations and the sequencing of multiple strains of the same species are aiding to study the role of recombination in bacterial evolution at the population level (Maiden, 2006). These analyses are revealing that most bacterial populations are far from being clonal and that recombination events can be detected by *in vitro* and *in silico* analyses (Perez-Losada *et al.*, 2006). The role of recombination in generating genetic variation and its contribution to overall genome plasticity is very different from one species to another. Estimates range from species in which recombination seems to be the major force of adaptive evolution (e.g. *Neisseria meningitidis*, *Streptococcus pneumoniae*, *Staphylococcus aureus*) to those that seem to evolve mainly by point mutations or other mechanisms, with recombination as a residual process (e.g. *Mycobacterium tuberculosis*, *Yersinia pestis*, *Bacillus anthracis*) (Feil and Spratt, 2001). However, most inferences on bacterial population structure are being reconsidered in the light of current analytical techniques that, as MLST, are changing dramatically long-held views such as the clonality of *Escherichia coli* populations (Wirth *et al.*, 2006). Nevertheless, these analyses have shown not only differences among taxa but also important differences within the corresponding genomes. Genes seem to have different recombination rates and, on top of that, highly recombining regions alternate with low recombining regions in bacterial genomes.

Nascent lineages that cannot (legitimately) recombine because of the accumulated divergence can acquire DNA from illegitimate sources. Instead of homologous recombination, other mechanisms can help to exchange genetic material among distant bacteria (Thomas and Nielsen, 2005). However, they are partly out of the scope of this review as we are interested in (1) reviewing mechanisms that prevent HGT between lineages, and (2) showing that the by-product of these mechanisms is the preferential sharing of DNA between evolutionarily close taxa.

Internal factors

Genomic architecture

Progress in the analyses of the structure of bacterial genomes, driven by the continuous sequencing of new genomes, is allowing their description and characterization beyond the cataloguing of genes included therein (Rocha, 2004). For example, the preferential distribution of genes in the leading strand has been usually explained by a selective pressure to avoid the collision of machineries in the leading strand during the simultaneous replication

and transcription processes in bacteria. This would allow a better replication and higher expression of genes (Brewer, 1988) but essentiality seems to be a more important factor, even constraining the number of possible rearrangements (Rocha and Danchin, 2003a,b). Recent studies have also identified codon usage domains in bacterial chromosomes that are important for the control of gene expression (Bailly-Bechet *et al.*, 2006). These results support the idea of domains in bacterial chromosomes larger than genes, operons or even 'über-operons' (Boccard *et al.*, 2005). The alteration of these macrodomains by rearrangements or the insertion of foreign sequences could have a detrimental effect on the fitness of the organism and, therefore, a 'compatibility filter' must operate in order to maintain chromosomal properties.

Such a filter has already been described by Lawrence and collaborators (Lawrence and Hendrickson, 2004; Hendrickson and Lawrence, 2006). These authors analysed the distribution of a certain kind of octamers along the bacterial chromosome. Repeats of these sequences are usually found in bacterial genomes in both strands; however, in some cases they are distributed asymmetrically and skewed towards the terminus of replication. This characteristic distribution seems to be associated to a specific role in the replication of bacteria; they are recognized by protein FtsK, whose function is related to translocation of the chromosome during cell division. The protein uses the accumulation of this repeat sequence as a signal for the location of the terminus of replication. Therefore, for a correct cell division it is necessary to maintain the skewed and asymmetrical distribution of these octamers. This possibility of selection at the level of chromosome structure could have important implications in bacterial chromosome dynamics. For instance, only inversions including the origin or terminus of replication might be evolutionarily feasible because other alternatives imply a disruption in the accumulation of these repetitive sequences and, in consequence, a disruption in the action of FtsK. Of more relevance in this context, DNA fragments compatible with the distribution of these repetitive units in a recipient genome will have more chances of being successfully transferred. Until now, different octamer sequences with similar distribution to the one described above have been identified and are usually shared by related taxa. This implies that successful transfers will be more likely between close, and therefore compatible, genomes.

Molecular mechanisms

The best-known mechanism to control the uptake of DNA in Bacteria is the methylation of a genome's own genetic material. DNA molecules without the specific methylation pattern are digested (Jeltsch, 2003). Recently one of these systems has been proposed as being responsible for foreign DNA control in *Staphylococcus aureus* (Waldron and Lindsay, 2006). These type I restriction-methylation systems require the action of three genes for the methylation process (*hsdR*, hsd*M*, *hsdS*). Variations in these genes throughout sequenced *S. aureus* genomes and in multistrain microarray analyses showed a significant correlation with the ten known dominant lineages of this species. Furthermore, the recognition of a sequence as alien by *S. aureus* lineages is based on different sequence profiles of these genes. The effect of this biased methylation is that (1) intra-lineage exchanges are more likely than inter-lineage exchanges and (2) there are preferential sharing links between different lineages. Obviously, this mechanism also acts upon sequences obtained from other species and therefore its presence results in a more likely exchange between *S. aureus* strains to the

point that it seems to be responsible for the identification of different lineages and the clonal structure of the species.

Another largely unresolved question about bacterial genome architecture is the characteristic GC/AT ratio (Bentley and Parkhill, 2004). Sometimes, lower GC contents are clearly associated with certain lifestyles, mainly in obligate intracellular pathogens and endosymbionts (Moran and Wernegreen, 2000). However, among the remaining genomes differences in GC content are remarkable and genus-specific. A possible role for these biases has been proposed. The protein H-NS has the ability to bind DNA, acting as a transcriptional regulator. Two independent experiments were able to determine that an important fraction of genes targeted by H-NS were low-GC compared with the average genome (Lucchini *et al.*, 2006; Navarre *et al.*, 2006). A low GC content together with the absence of these genes from the *Salmonella* core genome suggest that H-NS can act on HGT genes and repress their expression in a process called 'xenogenic' silencing. The silencing of these genes prevents their possible negative effects on the fitness of the genome. What is the fate of these silenced genes? Apparently bacteria have evolved mechanisms in order to selectively activate some of these genes (Navarre *et al.*, 2007). The fate of the remaining genes could be their pseudogenization. The H-NS mechanism in *Salmonella* suggests that transfers from donors with similar GC content maybe more likely to be retained in the genome because the H-NS mechanism is not at work. H-NS is a general regulator in bacteria with homologues mainly found in Gram-negative bacteria suggesting that silencing of recently acquired genes can be a general mechanism to control widespread horizontal gene transfer events (Tendeng and Bertin, 2003).

Functional association

Since the publication of the complexity hypothesis by Jain *et al.* (1999) it has been accepted that genes belonging to informational categories are less prone to HGT than operational ones. The hypothesis is based on two observations: (1) HGT has been continuous in the evolutionary history of bacteria, and (2) there are important differences between the degree of phylogenetic incongruence found in operational and metabolism-related genes. Jain *et al.* (1999) proposed that informational genes tend to be part of essential complexes interacting with many products whereas operational genes usually take part in simpler metabolic networks. The likelihood of integration and its effect on fitness should be very different, thus resulting in more successful transfers of operational genes than informational ones. Posterior studies with more genes and genomes have shed light on the importance of the distinction among operational and informational genes. Although there is clearly an association between the two factors it seems that the pattern is more complex than initially postulated and it is not only a by-product of the connectivity of the transferred gene.

In support of the complexity hypothesis, Nakamura *et al.* (2004) identified metabolism-related genes (cell surface, DNA binding and pathogenicity-related functions) as the more frequently transferred group and found very few cases of transfers involving informational genes. In addition, a more recent study with a very different approximation corroborated these observations (Pal *et al.*, 2005). These authors studied the influence of different factors on the stability of metabolic networks of bacteria. They identified several instances of horizontal gene transfer events and mapped them onto the metabolic network of *E. coli*. Their analyses revealed that (1) most genes involved in HGT belong to metabolism-related

categories, and (2) these genes usually occupy external nodes of the metabolic networks. These observations supported the complexity hypothesis because it is expected that genes attached to external nodes of a network carry out non-essential functions, which, in turn, are mainly encoded by operational genes. Therefore, the numbers of interactions in which a newly acquired gene is involved are fewer than those interactions needed when the transferred gene corresponds to a central node of the network.

However, other studies analysing the phenomenon from a phylogenetic point of view only agree partially with these observations. Phylogenetic analysis of incongruence has shown the presence of possible transfer events in an important proportion of core gene datasets (Bapteste *et al.*, 2005; Susko *et al.*, 2006). Most of the genes in these cores are included in informational categories. Also, recombination events in the elongation factor complex, even at large phylogenetic distances, have been reported (Inagaki *et al.*, 2006). A phylogenomic approach method called 'embedded quartets' has allowed inference of the number of transfer events among Cyanobacteria genomes and between Cyanobacteria and external species (Zhaxybayeva *et al.*, 2006). The analysis revealed these genomes as mosaics and the analysis of functional categories showed two distinct patterns. Transfers among cyanobacterial genomes showed no clear pattern, whereas those transfers that implied a non-cyanobacterial partner seemed to be biased towards operational genes, thus implying that transfers of informational genes are more likely between closely related genomes. The same pattern was observed in the screening of Xanthomonadales genomes in search of transfer events to the ancestor of the group and within the group (Comas *et al.*, 2006). Most ancestral transfers were from other Proteobacteria and showed no clear pattern of functional association although their mosaicism at a functional and genome scale prevented assigning a single phylogenetic origin to the group.

More recently the complexity hypothesis has been revisited and reformulated in light of new experimental data. The number of protein interactions seems to be a better predictor of the likelihood of transfer of a gene than the functional category it belongs to (Cohen *et al.*, 2011), a hypothesis supported by some experimental models of evolution (Omer *et al.*, 2010).

External factors

Ecological opportunity

The last stage in the genomic revolution, started with the publication of the *H. influenzae* genome in 1995 (Fleischman *et al.*, 1995), has been the study of microbial communities using metagenomic approaches. These methodologies are aimed at identifying not only the bacterial diversity in a specific niche but also the gene composition of its bacterial communities and the metabolic capabilities derived from them. There are two main factors to be considered in order to achieve these objectives. On the one hand, laboratory techniques have been developed that allow the fast isolation and sequencing of samples and on the other hand it is necessary to have a large enough taxon sampling of complete microbial genomes in order to assign sequences to their most closely related taxa.

In the last years, many studies have analysed the gene and species composition of very different environments. Some of them have revealed a low bacterial diversity with communities dominated by two or a few bacterial taxa, whereas other studies have identified niches in which bacterial diversity is much higher, with complex ecological networks relating them.

The number of new genes, those with no known function or not identified in any previously sequenced genome, is extraordinarily high even in communities composed by bacteria with clear relatives in the databases. All these studies are highlighting the vast amount of unknown microbial diversity and the limitations that this ignorance is imposing to our current views of bacterial evolution.

The analysis of environmental metagenomes is important in the context of HGT because one of the most important extrinsic factors affecting the process is ecological opportunity (see Chapter 4 for more details). Two genomes may be compatible in terms of genome architecture but if they occupy very different niches with no opportunity for contact then successful HGT events would be very unlikely. For example, Tyson *et al.* (2004) analysed the bacterial community structure of samples recovered from acid mine drainage (AMD) microbial biofilms. Only a few lineages dominated the community, mostly *Leptospirillum* species. The low-complexity of the environment seems to explain the presence of this dominant lineage and other low frequency Bacteria and Archaea (the *Ferroplasma* type II genome was sequenced from this study). In such an environment, the routes for DNA transfer are constrained. On the one hand, recombination seems to be the major evolutionary force driving speciation of the dominant *Leptospirillum* genomes. On the other hand, a few more cases of ancient HGT could be inferred including cases of inter-domain exchange. At the other extreme, the analysis of samples from the Sargasso Sea revealed a richer ecological structure composed by a range of 1800 to 48,000 genomic species (Venter *et al.*, 2004). Although almost all bacterial phyla were represented, most of the species proved to be related to already sequenced Proteobacteria genomes. Another remarkable fact was the presence of spatial patchiness in the frequency and distribution of these species and the identification of around 1.2 million new genes. Therefore, in such a complex community ecological opportunity for genetic exchange at almost all phylogenetic distances is possible in contrast to the AMD environment.

The two examples outlined above highlight also the need for new conceptual frameworks for the evolutionary genomic analysis of bacteria. The term 'pan-genome' has been proposed to describe all the genes present in any of the genomes from a species (Tettelin *et al.*, 2005). Species such as *Streptococcus agalactiae* have an open pan-genome and therefore it is expected that even the sequencing of 100 more strains will add new genes to it (Tettelin *et al.*, 2005). The analysis of closely related *Shigella*/*Escherichia* species shows similar results (Chen *et al.*, 2006). Other species, like *Bacillus*, seem to have a closed pan-genome: almost all genes that are part of one of its genomes are also present in the rest and the sequencing of further genomes will probably not add much to the known catalogue (Tettelin *et al.*, 2005). Analogously, a pan-genome for metagenomic studies could be outlining the so-called microbiome. An open pan-genome/microbiome clearly points to HGT as the main agent of evolutionary novelty. Focusing in the two metagenome examples given above, an environment such as the AMD dominated by *Leptospirillum* species is expected to have a closed microbiome while an environment such as that of the Sargasso Sea would have an unapproachable microbiome. Therefore, the opportunity for transfers is very different in these two environments, being much higher in the second case. It is also important to note that transfers across very large distances are not impossible as demonstrated by Tyson *et al.* (2004). However, as illustrated in this study, transfers from Bacteria to Archaea are usually related with adaptation to very specific niches, in which case the transfer of functions allowing survival in this niche is associated to a high fitness.

Phage genomics

Phage-mediated transfers are one of the possible mechanisms introducing foreign DNA in a bacterial genome. In the context of mechanisms and limitations for HGT, phages can be analysed from different points of view. They can be used as model organisms to study characteristics expected to be present in more complex organisms like Bacteria. Alternatively, they can be studied attending to their influence on bacterial chromosome evolution, as carriers of foreign genetic material.

Phages are clearly mosaic genomes resulting from multiple homologous and non-homologous recombination events. This mosaicism has been described both for double- and single-strand DNA phages (Lawrence *et al.*, 2002). However, the analysis of different T4-type phages has revealed limitations to rampant gene transfer (Filee *et al.*, 2006). These authors located two regions that seem not to be affected by HGT and whose main feature is the conservation of synteny across the analysed genomes. The genes composing these regions have coupled metabolic functions, characterized by many interactions. A disruption of these regions could result in important losses of fitness. As we have described, the conservation of genomic architecture seems to be an important factor for successful transfers in bacterial genomes. The evidence derived from phage genomics corroborates this importance.

Our current view of phages has changed from an early view as agents for the transportation of antibiotic resistance genes to consider them as catalysts of bacterial evolution (Canchaya *et al.*, 2003). Their lysogenic action used to be considered solely as the cause of conversion from a non-pathogen strain to a pathogen organism as shown by the phage encoded toxins of the causative agents of diphtheria or shigellosis (Brussow *et al.*, 2004). However, the ecology of phages also has to be taken into account. They have limited host ranges, mainly at the species level, although cases of broader ranges covering large phylogenetic distances have been described (Jensen *et al.*, 1998; Beumer and Robinson, 2005). It is known that bacterial strains use to share a common phage pool and therefore phage-mediated transfers are more likely between more closely related strains. The same example of the acid mine drainage used above is also useful to illustrate this point. The authors demonstrated that recombination between *Leptospirillum* species and between *Ferroplasma* species are phage dependent, as revealed by the analysis of their genomes (Tyson *et al.*, 2004). In another study, the analysis of *Prochlorococcus* genomes through a wide range of ocean light and depth conditions revealed that there are different ecotypes depending of these variables (Coleman *et al.*, 2006). The distinguishing feature of these ecotypes is the composition of different genomic islands in their genomes that have been introduced most likely by phages. These studies are pointing to phages as the main via of HGT in closely related microbial genomes as revealed by multistrain analyses of sequenced genomes (Brussow *et al.*, 2004; Aziz *et al.*, 2005). Finally, there is a dual role for phages in the transfer of genetic material. There are cases of phages with very restricted host ranges combined with cases of phages with a wide host spectrum and cosmopolitan distribution, which implies that they can act as a link between bacterial species and also across large geographic areas.

A proposal for Bacteria and Archaea

Our proposal for the nature and impact of gene transfer processes on microbial genomes is based on the idea that the likelihood of transfers decays with the phylogenetic distance between the lineages involved, in analogy to the lower recombination rate between more

divergent sequences. We have explained this effect as a by-product of different mechanisms that allow bacteria to prevent the incorporation, silence or eliminate foreign material from more or less distant sources. Some of these mechanisms help to define intra-species lineages whereas others prevent inter-species exchange and, presumably, other mechanisms that surely act at other phylogenetic depths have still to be identified. However, horizontal gene transfer events between distant taxa are still detectable. Among them only those transfers from distant sources with a high, positive effect on the fitness of the recipient will be retained. Usually, those successful transfers are dependent on environmental conditions like the exchange of niche-specific genes between Bacteria and Archaea that share the same extreme environments (Mongodin *et al.*, 2005).

Although the decay of recombination rates with divergence has been demonstrated experimentally (Majewski and Cohan, 1998), only now the growing number of sequenced genomes is allowing to study the age and number of transfers between species. There is a growing body of evidence supporting the assumption that the likelihood of horizontal gene transfers decays with time since divergence from the common ancestor (Andam and Gogarten, 2011; Popa *et al.*, 2011). As mentioned previously, the evolution of Xanthomonadales has been driven by ancient and recent HGT events from other Proteobacteria and from other Xanthomonadales genomes respectively (Comas *et al.*, 2006). We tested more properly the hypothesis of the decay of HGT events with distance using the genome of the opportunistic pathogen *Legionella pneumophila*. As a difference with other genome-wide studies of horizontal gene transfer events we decided to focus on an single species and tested its likelihood of receiving genes from other bacterial groups after removal of those gene families with poor phylogenetic signal. The approach allowed us, firstly, to test for the number of events occurred between *Legionella* and other groups of bacteria and, secondly, to differentiate between noise and signal when doing phylogenetic assignments. In that way we have shown for the first time in Bacteria a high correlation between phylogenetic distance and the number of horizontal gene transfer events to a specific genome (Coscollá *et al.*, 2011). Although not explicitly tested as in our work with *L. pneumophila*, other studies pointing towards similar results have been published. The Cyanobacteria genomes analysed by Zhaxybayeva *et al.* (2006) revealed that around 50% of their genes have suffered at least one HGT event. What is more relevant for our proposal, intra-Cyanobacteria transfers were more frequent than between Cyanobacteria and external genomes. Other studies have shown similar patterns in Corynebacteria (Marri *et al.*, 2007). In addition, analyses of multiple genomes covering large phylogenetic distances point towards the hypothesis of phylogenetic preferential sharing. In the most complete analysis published up to now, Beiko and collaborators (2005) screened 144 sequenced genomes including Bacteria and Archaea. They derived a supertree from the single-gene phylogenies obtained from all the gene families retrieved from these genomes and then tried to reconcile every gene tree and the supertree. As the reconciliation of a gene tree equals to a putative HGT event, their conclusion was that most of the inferred transfers were between genomes of the same group, for example between genomes of Gammaproteobacteria. Therefore the analysis showed as preferential partners in HGT events those of closely related genomes, as predicted by Gogarten *et al.* (2002).

To simplify our argument we will consider a genome whose capability to accept new genetic material is constant though time. In this case, a linear income of foreign material during the evolutionary history of the genome is expected, as shown in Fig. 5.2. Therefore, the current genome derived from this lineage will present ancient, recent and ongoing

HGTs. From a methodological point of view the detection of these events requires very different approximations.

Ongoing transfers are composed mainly by homologous recombination within the lineage. This kind of transfers is only detectable in a population genetics framework that allows inferring events and rates of recombination. There are many examples on how to use these techniques in bacterial populations (Perez-Losada *et al.*, 2006). From a phylogenomic point of view, there is no effect on genome phylogenies not only because there is not enough phylogenetic sampling but also because at this scale recombination acts synergistically with the vertical signal as a cohesive force of the group (Fraser *et al.*, 2007).

However, homologous recombination is not the only mechanism for acquiring foreign material. Recent transfers from other genomes can also be detected by surrogate methods (Lawrence and Ochman, 2002). These methods identify atypical regions in the genomes based on a significant difference in some compositional measure. Each genome is under different mutational pressures (Sueoka, 1993), which result in specific patterns of, for example, nucleotide composition (Lawrence and Ochman, 1997; Lawrence and Ochman, 1998; Ochman *et al.*, 2000), codon usage bias (Medigue *et al.*, 1991), dinucleotide frequencies (Karlin and Burge, 1995; Karlin, 2001) and sequence patterns detected by Markov models (Hayes and Borodovsky, 1998). In consequence, the introduction of new DNA by an HGT event results in the integration in the genome of a sequence with different features than those typical of the recipient genome (Ochman *et al.*, 2000). Usually each of these measures tends to generate different sets of atypical genes although it has been argued that the reason is that each one tests a different hypothesis, being best suited to detect transfers from different genomes (Azad and Lawrence, 2005). Although not all the atypical genes detected result from HGTs, they are a good measure of the recent and ongoing transfers in a genome. However, the main limitation of these methods is the reduction and eventual elimination of atypical signals with time (Lawrence and Ochman, 1997). This process, called amelioration, starts when the DNA integrates in the genome and therefore when the sequence starts to undergo the same biases than the rest of the genome. The process of amelioration is relatively fast and therefore the power to detect atypical genes decreases with time, being valid only for recent transfers. The main advantage of this approach is that it does not rely on a comparative analysis and therefore those genes excluded from a phylogenetic analysis owing to their presence in only one or a few lineages can be studied. Phylogenetic methods will detect these transfers if the divergence between the donor and receptor is large enough. For example, a recent transfer between Bacteria and Archaea is easy to detect whereas a recent transfer between *Salmonella typhimurium* and *Escherichia coli* will be more difficult and would be only detected if the number of strains/species used in the study is large enough to reveal the donor and recipient genomes.

Older transfers can be detected only by phylogenetic methods and their detection depends to a large extent on the time elapsed since the transfer. If the gene was transferred relatively recently, then it would show a clearly discordant phylogenetic signal, grouping the receptor in the donor's clade. However, if this transfer was older then the gene would carry two opposite signals: that from the donor, originated before the transfer, and the signal from the receptor, originated after the transfer. The first signal would suffer a process of phylogenetic amelioration, which is partly due to the compositional amelioration and to sharing selective pressures with related genomes. For example, positive selection could favour the same non-synonymous changes in the genes of a clade of genomes. Usually this translates

into a lowly supported assignment of these genes to any clade. When the transfer is very ancient it is often impossible to differentiate between a transferred gene and the genes of the receptor genome because most of the phylogenetic history that could be inferred is shared with the rest of the related genomes. The deeper the transfer the more difficult it would be to detect it. This also explains the presence of noise in terms of horizontal signal. Those gene transfers that are old, but not very ancient, will show support both for the donor phylogeny and for the receptor phylogeny. Therefore, some noise, such that one is unable to differentiate between the assumed species tree and the transfer hypotheses, is the hallmark of an old transfer event and is revealing a mixture of donor and receptor signals.

According to this model, the lineage considered will be receiving foreign DNA at a more or less constant rate, depending on its environment, as shown in Fig. 5.2. Next, the interplay between selection and effective population sizes will act filtering each transfer event, which will lead to differential incorporation of foreign material over time. Particularly, selection will act so that newly acquired material will be incorporated either when it has almost no impact on fitness (neutral or quasi-neutral effect which will depend on the effective population size) or when it represents a strong fitness advantage as in the case of new niche occupation or acquisition of new capabilities to compete with related strains. A free-exchange scenario, in which a lineage has constant access to foreign material through time is an ideal situation that might fit highly promiscuous bacteria like Xanthomonadales, which seem to lack most selective filters that prevent gene exchange. Other bacteria could have different behaviours. For example, a free-living bacterial lineage could receive a gene island that eventually allows its adaptation to a new niche. During the adaptation process population sizes, recombination rates and possibilities of gene transfers usually reduce, thus lowering the transfer rate. However, as the lineage becomes adapted to the new environment, diversification by periodic selection episodes and the acquisition of niche-specific genes could lead to an increase in the number and diversity of bacterial lineages or ecotypes. These could differentiate up to a point in which recombination between them is prevented although non-homologous transfers continue (Fig. 5.2).

This seems to be the case for the distribution of *Prochlorococcus* ecotypes in the oceans (Johnson *et al.*, 2006). Despite sharing a core genome, these ecotypes present a very large variability in the remaining gene contents, which reside mostly in genomic islands. A high correlation between genomic island composition and structure is found with the niche each ecotype occupies. These islands have been acquired by phage-mediated transfers that, in consequence, acted as drivers of *Procholorococcus* diversification. The establishment of the lineage in a new niche allows new opportunities for recombination and gene sharing.

But the acquisition of genes that promote the invasion of a new niche could also imply that some new or old genetic material is no longer necessary because the organism has specialized in exploiting the new niche. This is the case for most pathogens derived from a free-living ancestor. The free-living genome receives transfers from other genomes. Eventually some of these transfers, presumably pathogenicity islands, allow it to infect a host as a pathogen. As the transition to a pathogenic lifestyle continues the capacity for receiving external genetic material decreases both because it is less important and because the environment offers fewer opportunities. Most bacterial pathogens show an intermediate pattern of reduced HGT capacity and genome decay signals due to constraints both in population sizes and in the selective advantages of new genetic material (Fig. 5.2). The most extreme cases are those of intracellular pathogens and endosymbionts, which couple specialization

in the new host with loss of non-essential genes for their lifestyle. Intracellular pathogens in which these phenomena have been described are *Mycobacterium leprae* (Cole *et al.*, 2001) or *Ricketssia prowazeckii* (Müller and Martin, 1999) (Fig. 5.2). The genomes of bacterial endosymbionts undergo the same processes as obligatory intracellular pathogens but much more accentuated, with perfect gene-order conservation among strains and with no opportunity for gene exchange (Wernegreen, 2002). In these cases, horizontal gene transfers to the non-intracellular ancestors of these genomes might be detected but only if they have not been lost along the genome reduction process.

Conclusion

It has been estimated that every gene family has suffered at least one transfer event during its evolutionary history (Dagan and Martin, 2007). The impact of many aspects of horizontal gene transfer is still hotly debated (Ragan and Beiko, 2009) but currently it is clear that HGT is a major force shaping the genome contents of prokaryotes. From the initial reports showing the surprisingly high proportion of foreign genetic material in bacterial genomes we have now moved to quantitative approaches revealing the complex interplay between gene loss, horizontal gene transfer and prokaryotes' lifestyles. We think that the combination of shotgun sequences generated over the last 15 years with the large amount of information generated by the new sequencing technologies will give us insights into the horizontal gene transfer process to a scale and detail that will solve many open issues about its past, present and future impact on bacterial innovation.

References

Andam, C.P. and Gogarten, J.P. (2011). Biased gene transfer in microbial evolution. Nat. Rev. Micro. *9*, 543–555.

Azad, R.K. and Lawrence, J.G. (2005). Use of artificial genomes in assessing methods for atypical gene detection. PLoS Comput. Biol. *1*, e56.

Aziz, R.K., Edwards, R.A., Taylor, W.W., Low, D.E., McGeer, A., and Kotb, M. (2005). Mosaic prophages with horizontally acquired genes account for the emergence and diversification of the globally disseminated m1t1 clone of *Streptococcus pyogenes*. J. Bacteriol. *187*, 3311–3318.

Bailly-Bechet, M., Danchin, A., Iqbal, M., Marsili, M., and Vergassola, M. (2006). Codon usage domains over bacterial chromosomes. PLoS Comput. Biol.*2*, e37.

Bapteste, E., Susko, E., Leigh, J., MacLeod, D., Charlebois, R.L., and Doolittle, W.F. (2005). Do orthologous gene phylogenies really support tree-thinking? BMC Evol. Biol. *5*, 33.

Beiko, R.G., Harlow, T.J., and Ragan, M.A. (2005). Highways of gene sharing in prokaryotes. Proc. Natl. Acad. Sci. U.S.A. *102*, 14332–14337.

Bentley, S.D. and Parkhill, J. (2004). Comparative genomic structure of prokaryotes. Annu. Rev. Genet. *38*, 771–791.

Beumer, A. and Robinson, J.B. (2005). A broad-host-range, generalized transducing phage (SN-T) acquires 16S rRNA genes from different genera of bacteria. Appl. Environ. Microbiol. *71*, 8301–8304.

Boccard, F., Esnault, E., and Valens, M. (2005). Spatial arrangement and macrodomain organization of bacterial chromosomes. Mol. Microbiol. *57*, 9–16.

Brewer, B.J. (1988). When polymerases collide: Replication and the transcriptional organization of the *E. coli* chromosome. Cell *53*, 679–686.

Brussow, H., Canchaya, C., and Hardt, W.D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol. Mol. Biol. Rev. *68*, 560–602.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L., and Brussow, H. (2003). Phage as agents of lateral gene transfer. Curr. Opin. Microbiol. *6*, 417–424.

Chen, S.L., Hung, C.S., Xu, J., Reigstad, C.S., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R.R., Ozersky, P., *et al.* (2006). Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. Proc. Natl. Acad. Sci. U.S.A. *103*, 5977–5982.

Cohan, F.M. (2002). Sexual isolation and speciation in bacteria. Genetica *116*, 359–370.

Cohen, O., Gophna, U., and Pupko, T. (2011). The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. Mol. Biol. Evol. *28*, 1481–1489.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D., *et al.* (2001). Massive gene decay in the leprosy bacillus. Nature *409*, 1007–1011.

Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., DeLong, E.F., and Chisholm, S.W. (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. Science *311*, 1768–1770.

Comas, I., Moya, A., Azad, R.K., Lawrence, J.G., and González-Candelas, F. (2006). The evolutionary origin of xanthomonadales genomes and the nature of the horizontal gene transfer process. Mol. Biol. Evol. *23*, 2049–2057.

Coscollá, M., Comas, I., and González-Candelas, F. (2011). Quantifying nonvertical inheritance in the evolution of *Legionella pneumophila*. Mol. Biol. Evol. *28*, 985–1001.

Dagan, T. and Martin, W. (2007). Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. Proc. Natl. Acad. Sci. U.S.A. 0606318104.

Feil, E.J. and Spratt, B.G. (2001). Recombination and the population structures of bacterial pathogens. Annu. Rev. Microbiol. *55*, 561–590.

Filee, J., Bapteste, E., Susko, E., and Krisch, H.M. (2006). A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. Mol. Biol. Evol. *23*, 1688–1696.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., *et al.* (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science *269*, 496–512.
<AU: Please provide 10 author names before et al>

Fraser, C., Hanage, W.P., and Spratt, B.G. (2007). Recombination and the nature of bacterial speciation. Science *315*, 476–480.

Gogarten, J.P., Doolittle, W.F., and Lawrence, J.G. (2002). Prokaryotic evolution in light of gene transfer. Mol. Biol. Evol. *19*, 2226–2238.

Hayes, W.S. and Borodovsky, M. (1998). How to interpret an anonymous bacterial genome: machine learning approach to gene identification. Genome Res. *8*, 1154–1171.

Hendrickson, H. and Lawrence, J.G. (2006). Selection for chromosome architecture in bacteria. J. Mol. Evol. *62*, 615–629.

Inagaki, Y., Susko, E., and Roger, A.J. (2006). Recombination between elongation factor 1α genes from distantly related archaeal lineages. Proc. Natl. Acad. Sci. U.S.A. *103*, 4528–4533.

Jain, R., Rivera, M.C., and Lake, J.A. (1999). Horizontal gene transfer among genomes: The complexity hypothesis. Proc. Natl. Acad. Sci. U.S.A. *96*, 3801–3806.

Jeltsch, A. (2003). Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? Gene *317*, 13–16.

Jensen, E.C., Schrader, H.S., Rieland, B., Thompson, T.L., Lee, K.W., Nickerson, K.W., and Kokjohn, T.A. (1998). Prevalence of broad-host-range lytic bacteriophages of *Sphaerotilus natans*, *Escherichia coli*, and *Pseudomonas aeruginosa*. Appl. Environ. Microbiol. *64*, 575–580.

Karlin, S. (2001). Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. Trends Microbiol. *9*, 335–343.

Karlin, S. and Burge, C. (1995). Dinucleotide relative abundance extremes: a genomic signature. Trends Genet.. *11*, 283–290.

Lawrence, J.G. and Hendrickson, H. (2004). Chromosome structure and constraints on lateral gene transfer. Dynam. Genet. 319–336.

Lawrence, J.G. and Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. *44*, 383–397.

Lawrence, J.G. and Ochman, H. (1998). Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. U.S.A. *95*, 9413–9417.

Lawrence, J.G. and Ochman, H. (2002). Reconciling the many faces of lateral gene transfer. Trends Microbiol. *10*, 1–4.

Lawrence, J.G., Hatfull, G.F., and Hendrix, R.W. (2002). Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J. Bacteriol. *184*, 4891–4905.

Lucchini, S., Rowley, G., Goldberg, M.D., Hurd, D., Harrison, M., and Hinton, J.C.D. (2006). H-NS mediates the silencing of laterally acquired genes in Bacteria. PLoS Pathog. *2*, e81.
Maiden, M.C.J. (2006). Multilocus sequence typing of Bacteria. Annu. Rev. Microbiol. *60*.
Majewski, J. and Cohan, F.M. (1998). The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. Genetics *148*, 13–18.
Marri, P., Hao, W., and Golding, G.B. (2007). The role of laterally transferred genes in adaptive evolution. BMC Evol. Biol. *7*, S8.
Medigue, C., Rouxel, T., Vigier, P., Henaut, A., and Danchin, A. (1991). Evidence for horizontal gene transfer in *Escherichia coli* speciation. J. Mol. Biol. *222*, 851–856.
Molmeret, M., Horn, M., Wagner, M., Santic, M., and bu Kwaik, Y. (2005). Amoebae as training grounds for intracellular bacterial pathogens. Appl. Environ. Microbiol. *71*, 20–28.
Mongodin, E.F., Nelson, K.E., Daugherty, S., DeBoy, R.T., Wister, J., Khouri, H., Weidman, J., Walsh, D.A., Papke, R.T., Sanchez Perez, G., *et al.* (2005). The genome of *Salinibacter ruber*: Convergence and gene exchange among hyperhalophilic bacteria and archaea. Proc. Natl. Acad. Sci. U.S.A. *102*, 18147–18152.
Moran, N.A. and Wernegreen, J.J. (2000). Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends Ecol. Evol. *15*, 321–326.
Müller, M. and Martin, W. (1999). The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes. Bioessays *21*, 377–381.
Nakamura, Y., Itoh, T., Matsuda, H., and Gojobori, T. (2004). Biased biological functions of horizontally transferred genes in prokaryotic genomes. Nat. Genet. *36*, 760–766.
Navarre, W.W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S.J., and Fang, F.C. (2006). Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. Science *313*, 236–238.
Navarre, W.W., McClelland, M., Libby, S.J., and Fang, F.C. (2007). Silencing of xenogeneic DNA by H-NSG-facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA. Gene. Dev. *21*, 1456–1471.
Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000). Lateral gene transfer and the nature of bacterial innovation. Nature *405*, 299–304.
Omer, S., Kovacs, A., Mazor, Y., and Gophna, U. (2010). Integration of a foreign gene into a native complex does not impair fitness in an experimental model of lateral gene transfer. Mol. Biol. Evol. *27*, 2441–2445.
Orskov, F. and Orskov, I. (1983). From the national institutes of health. Summary of a workshop on the clone concept in the epidemiology, taxonomy, and evolution of the Enterobacteriaceae and other bacteria. J. Infect. Dis. *148*, 346–357.
Pal, C., Papp, B., and Lercher, M.J. (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat. Genet. *37*, 1372–1375.
Perez-Losada, M., Browne, E.B., Madsen, A., Wirth, T., Viscidi, R.P., and Crandall, K.A. (2006). Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. Infect. Genet. Evol. *6*, 97–112.
Popa, O., Hazkani-Covo, E., Landan, G., Martin, W., and Dagan, T. (2011). Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. Genome Res. *21*, 599–609.
Ragan, M.A. and Beiko, R.G. (2009). Lateral genetic transfer: open issues. Phil. Trans. R. Soc. B *364*, 2241–2251.
Rocha, E.P.C. (2004). The replication-related organization of bacterial genomes. Microbiology *150*, 1609–1627.
Rocha, E.P.C. and Danchin, A. (2003a). Essentiality, not expressiveness, drives gene-strand bias in bacteria. Nat. Genet. *34*, 377–378.
Rocha, E.P.C. and Danchin, A. (2003b). Gene essentiality determines chromosome organisation in bacteria. Nucl. Acids. Res. *31*, 6570–6577.
Stenderup, J. and Orskov, F. (1983). The clonal nature of enteropathogenic *Escherichia coli* strains. J. Infect. Dis. *148*, 1019–1024.
Sueoka, N. (1993). Directional mutation pressure, mutator mutations, and dynamics of molecular evolution. J. Mol. Evol. *37*, 137–153.
Susko, E., Leigh, J., Doolittle, W.F., and Bapteste, E. (2006). Visualizing and assessing phylogenetic congruence of core gene sets: a case study of the gamma-proteobacteria. Mol. Biol. Evol. *23*, 1019–1030.
Tendeng, C. and Bertin, P.N. (2003). H-NS in Gram-negative bacteria: a family of multifaceted proteins. Trends Microbiol. *11*, 511–518.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., *et al.* (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial 'pan-genome'. Proc. Natl. Acad. Sci. U.S.A. *102*, 13950–13955.

Thomas, C.M. and Nielsen, K.M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat. Rev. Micro. *3*, 711–721.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J.F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature *428*, 37–43.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science *304*, 66–74.

Waldron, D.E. and Lindsay, J.A. (2006). Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between *S. aureus* isolates of different lineages. J. Bacteriol. *188*, 5578–5585.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L.H., Karch, H., Reeves, P.R., Maiden, M.C.J., Ochman, H., *et al.* (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol. Microbiol. *60*, 1136–1151.

Zhaxybayeva, O., Gogarten, J.P., Charlebois, R.L., Doolittle, W.F., and Papke, R.T. (2006). Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. Genome Res. *16*, 1099–1108.